DOCUMENT RESUME

ED 345 741 IR 054 044

AUTHOR	Rutimann, Hans; Lynn, M. Stuart
TITLE	Computerization Project of the Archivo General de Indias, Seville, Spain. A Report to the Commission on Preservation and Access.
INSTITUTION	Commission on Preservation and Access, Washington, DC.
PUB DATE	Mar 92
NOTE	20p.
AVAILABLE FROM	Commission on Preservation and Access, 1400 16th Street, N.W., Suite 740, Washington, DC 20036-2217 ($5.00 prepaid).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; *Archives; Bibliographic Records; Computer Graphics; *Computer System Design; Database Management Systems; Foreign Countries; History; Imagery; Information Dissemination; Information Networks; Local Area Networks; *Preservation; *Records Management
IDENTIFIERS	Columbus (Christopher); Multimedia Materials; North America; South America; *Spanish History; Workstations

ABSTRACT

The Archivo General de Indias is operating a massive project to preserve and make accessible the contents of the 45 million documents and 7,000 maps and blueprints comprising the written heritage of Spain's 400 years in power in the Americas. The current objective is to scan about 10 percent of the archive (or about 8 million images) in preparation for the 1992 Seville World's Fair and the Columbus quincentenary. The archive was established by Carlos III in 1785 to collect in one place all documents associated with the Spanish colonization of the Americas, and documents date from the 15th through the 19th centuries, inclusive. It is expected that the King of Spain, who is very interested in the project, will officially open the "electronic archive" at the beginning of the quincentenary celebrations. This report provides information on the progress of the project; support being provided by the Ministry of Culture, IBM Spain, the Foundation Areces, and the Archivo itself; the bibliographic database; selection criteria; and the components of the system, i.e., an image database, a bibliographic database, and a management database linked together by an IBM token-ring local area network (LAN). The scanning of images and workstation access are also described. The record structure of the bibliographic database and examples of forms completed by archivists are appended. (Author/MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The Commission on Preservation and Access

ED 345741

U.S. DEPARTMENT OF EDUCATION
(Office of Educational Research and Improvement)
(EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC))

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

COMPUTERIZATION PROJECT OF THE ARCHIVO GENERAL DE INDIAS, SEVILLE, SPAIN

A Report to the Commission on Preservation and Access

by
Hans Rütimann, Consultant, International Project
and
M. Stuart Lynn, Member, Technology Assessment Advisory Committee,
and Vice President, Information Technologies, Cornell University

March 1992

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Maxine K. Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

1400 16th Street, N.W., Suite 740, Washington, D.C. 20036-2217 • (202) 939-3400

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

The Commission is accessible via BITNET (SITTS@GWUVM) and by FAX (202) 939-3407.

BEST COPY AVAILABLE

COMPUTERIZATION PROJECT OF THE ARCHIVO GENERAL DE INDIAS, SEVILLE, SPAIN

This report to the Commission on Preservation and Access was prepared by Hans Rütimann, International Project Consultant, and M. Stuart Lynn, a member of the Technology Assessment Advisory Committee, after visits to the Archivo General de Indias in 1991. The Commission sponsored the inquiries into this project to learn more about the technical and operational implications of large-scale image scanning. Rütimann visited the facility in late April-early May 1991, and the Commission then asked Lynn to assess the project's technical aspects in September 1991.

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2217

March 1992

Reports issued by the Commission on Preservation and Access are intended to stimulate thought and discussion. They do not necessarily reflect the individual views of Commission members.

Additional copies are available for $5.00 from the above address. Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access." Payment must be in U.S. funds; do not send cash.

This publication has been submitted to the ERIC Clearinghouse on Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

COPYRIGHT 1992 by The Commission on Preservation and Access. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction for noncommercial purposes, including educational advancement, private study, or research will be granted. Full credit must be given to both the author(s) and The Commission on Preservation and Access.

The Archivo General de Indias is operating a massive project to preserve and make accessible the contents of the 45 million documents and 7,000 maps and blueprints comprising the written heritage of Spain's 400 years in power in the Americas. The present objective is to scan about 10 percent of the archivo (or about eight million images) in preparation for the 1992 Seville World's Fair and the Columbus quincentenary. The Archivo was established by Carlos III in 1785 to collect in one place all documents associated with the Spanish colonization of the Americas. Documents date from the 15th through 19th centuries, inclusive. It is expected that the King of Spain, who is very interested in the project, will officially open the "electronic archive" at the beginning of the quincentenary celebrations.

THE PROJECT

Four institutions are involved: The Ministry of Culture, IBM Spain, the Foundation Ramón Areces -- all in Madrid -- and the Archivo itself in Seville. Basically, the project consists of three parts: an image database, a bibliographic database, and an archive management system. The technical development work is carried out at the IBM Scientific Center located at the Universidad Autónoma north of Madrid, and all of the cataloging and scanning is done at the Archivo. The final system will be installed at the Archivo.

The project involves more than 100 people and is directed by a Coordinating Committee composed of Jorge Semprún Maura, the Minister of Culture; Fernando de Asúa Alvarez, Chairman, IBM Spain; and Ramón Areces Rodríguez, Chairman of the Ramón Areces Foundation and son of the founder Don Ramón Areces, who died two years ago. Project management, however, is under the "Direction Committee" composed of Pedro González, Director del Centro de Información Documental de Archivo, Ministry of Culture; Juan P. Secilla, Director, IBM Scientific Center, Madrid; and Rafael Ramírez, Ramón Areces Foundation.

The first phase, launched in 1986 and scheduled for completion by March, 1992, is estimated to cost a total of $10 million. Most of the funding is provided by the Foundation; the equipment and some staff are supplied by IBM Spain.

Even though the goals for the first phase have been scaled down from original newspaper accounts, it is still an awesome undertaking. There are 45 million documents in

the Archivo. Since all documents are written on front and back, the first phase's goal is to scan and control bibliographically about eight million pages plus maps and other prints.

There are about 43,000 "Bundles" ("Legajos") of documents in the Archivo comprising about 80-86 million pages. There are also about 8,000 maps and plans, most of which are in color. A bundle is a logical collection of documents, tied together with a ribbon and stored in a cardboard box roughly the size of a typical office filing box (about 11"x14"x4"). These bundles are stored upright and side by side along the shelves of the Archivo.

The digital scanning of the documents does not, of course, preserve them. These documents are probably sufficiently important to require conservation treatment. Scanning, however, does contribute to conservation insofar as it reduces the handling of the documents and resulting wear and tear.

The Ministry of Culture

Pedro Gonzáles was the guide in Madrid for both visits. He is an archivist, his title is Director del Centro de Información Documental de Archivo, and he works for the Dirección General de Bellas Artes y Archivo. His department had been working on the organization of the Presidential State Archives when it was given responsibility for the Seville project. Even though González is very interested and knowledgeable about all aspects of the project, he leaves no doubt that the Ministry's concern centers on the development of systems to manage archives. The plan is to use the system being developed by the Seville Archivo to other state archives thoughout Spain. The bibliographic database probably will be made available statewide through the network of the Ministry of Culture (Puntos de Información Cultural) and the national academic network, IRIS, which is linked to the Internet. Historians hope that the database will give clues to the development of bureaucratic systems in the modern world and will help them trace the intricacies of Spain's growth of power in the Americas.

Originally the project's goal was to catalog and scan only documents from the Archivo General de Indias itself. Recently, however, a decision was made to round out certain subcollections by scanning a number of documents from the Archivo Histórico Nacional and the Archivo de Simancas. Apparently, there are important documents in these archives pertaining to the Americas, such as military orders. Some documents will be temporarily transferred to Seville to be scanned; others will be scanned directly in Madrid and Simancas as soon as scanning capabilities are established there. It is also the general intention, if funding permits, to expand the use of the overall technology as the basis for digital preservation and access in all major archives in Spain. Mr. González summarized the project's goals as follows:

-- Preservation and diffusion of the historical heritage;
-- Application of new technology to historical archives;
-- Enhancement of the legibility of documents;
-- Support for the activities commemorating the quincentenary of the Americas' discovery.

IBM Spain

IBM and the Foundation are both contributing technical personnel to the project. Together, they are designing and implementing the overall system (see Technical Information, below) and conducting research on image compression and filtering techniques pertaining to old manuscripts.

IBM contributes staff, hardware and software. At the Universidad Autónoma about ten miles from Madrid, programmers and systems analysts are working on the second prototype bibliographic system for data retrieval, linkage of the text and image databases, and the user-oriented manipulation of images. Demonstrations of the prototype were impressive. The indexing system is detailed, contextual, and based on meticulous preparation of the data sheets by bibliographers in Seville. It was possible for a user of the system to retrieve the names of persons in a small Spanish village who applied for permission to emigrate to Chile some 400 years ago. In addition, it was possible to summon the appropriate document to the monitor to view the original application.

Available for viewing at IBM Spain was the display of a document, with a simultaneous display on another monitor explaining the document, for example, "Treaty of Tordesillas, signed by the Catholic Monarch and King John II of Portugal, on the demarcation and boundaries of the ocean; Portuguese version of the Treaty (June 7, 1494)." The rendering of the image was clear and every word could be seen.

Parts of the document or all of it can easily be enhanced by blocking out spots of ink bleeding through from the reverso. The document can be rotated on the screen, a very useful feature since much writing is on the margins and extends in all directions. IBM staff were optimistic that the second prototype of the entire system would be ready by the end of 1991. IBM staff represent the magnitude of the project as follows:

-- 80-86 million pages contained in 43,000 bundles of documents.

-- 8,000 maps and plans; 25,000 pages of inventories and catalogues, to be used by 15,000 researchers per year.

-- Remote online access is viewed as unlikely because of communications problems and a weak communications infrastructure.

-- Local access via CDs and optical disks is considered an option, but no long-term plans for wider dissemination of either the bibliographic or the image database have been made so far.

Foundation Ramón Areces

Don Ramón, as he is referred to by Rosario Parra Calla, the Director of the Archivo, was quite intrigued when the possibility of "digitizing" the Archivo was first discussed. Unfortunately, he died two years ago and will not witness his Foundation's remarkable achievements. Ramón Areces, a Cuban, had a life-long interest in American history,

particularly at a time when it was inextricably linked with Spain's. A self-made man, he settled in Spain, started a small store, and parlayed his business into a nationwide chain of department stores, El Corte Inglés.

The connections among the key players in this project become clearer when one learns that El Corte Inglés is IBM's largest customer in Spain. Also, as one of the largest computer users, the Foundation Ramón Areces can draw not only on its financial assets, but also on the technical expertise of the department stores' large technical staff. In fact, at any given time since the project's beginnings in 1986, between 12 and 15 El Corte Inglés employees have been working on the Archivo project fulltime.

The Archivo

At a time when Spanish control of the Americas was already weakening, Carlos III believed that a magnificent edifice, housing the entire written record of the colonies, would become a symbol of strength and consolidate Spain's enduring claim to the new world.

In July 1779, Carlos III ordered the scholar and "Cosmógrafo Mayor de Indias" Juan Bautista Muñoz to write a history of the New World. For several years Muñoz worked in the central archives of Simancas, organizing and cataloguing documents. At the same time, he and Carlos III were looking for an appropriate building where all documents could be archived in one place.

"The country's most suited building" was found in Seville, in the former commodity market next to the cathedral. The structure was originally built at the end of the sixteenth century to get the traders out of the cathedral. At the time, to renovate an old palace in order to serve an entirely new purpose was a daring concept. It still is, and that is precisely what is happening today, with high-tech equipment installed in the midst of rows upon rows of neatly bundled documents dating back centuries.

Bundles are packed high on stacks, some of which are on moving tracks. Each bundle is tied with white tape and encased in hard covers. The quality of paper is excellent, even though the watermarks indicate that the paper was produced some 400-500 years ago. Some documents are damaged, most frequently by waterstains and holes caused by acidic ink. The documents are in signatures of varying number, from a single page to as many as fifty. One single-page document is an application for travel to Peru, dated 1493 (granted).

For the computerization project, bibliographers work with the bundles, filling in data sheets for the bibliographic database. In another room, the data is keyboarded and a floppy disk accompanies the bundles to the large scanning room. Fifteen scanners are in two-shift operation; the scanning room sounds like a beehive or a nest of angry hornets because of the equipment's high-pitched whine. As the operator puts sheet after sheet on the scanner, the accompanying text is displayed on a monitor. By adding a control number from the bibliographic database to the scanned image, the link between the two databases is established. The sight of this high-tech operation in a converted baroque hall is indelible.

All materials in color, primarily maps, are first microfilmed by a service bureau in Madrid using Cibachrome. Then the fiche is digitized and, upon request, prints are produced off the fiche. According to Pedro González, the color quality of the prints is still

unsatisfactory, and he hopes to improve on it. As a side-product of the scanning process, then, a valuable color microfiche collection becomes available. A print service is planned for both the bibliographic information and images.

The Bibliographic Database

Even though the indexing system is impressive and well planned for the specific retrieval needs of the Archivo, there are no current plans for wider dissemination of the data, i.e., sharing it with bibliographic utilities in other countries and continents.

The record structure (see Appendix) reflects its local context, starting with tag 000 under "Información de control," more tags under the heading of "Información básica," and still more under "Información descriptiva." The tagging scheme is not related to available national or international standards.

By and of itself, this is not too tragic. Every expert who saw the tagging scheme agreed that it could be translated to fit some form of commonly known record structure. However, the creation of a totally new bibliographic record structure for a massive archival project in this day and age indicates that the Seville project was conceived, planned, and carried out as a regional project -- to better serve the needs of researchers who undertake the trek to Seville. More information about the database can be found in the Technical Information section.

Selection Criteria

As mentioned earlier, 10% of the archives have initially been selected for digitization. The primary criterion used to select documents for inclusion in the Image Database is frequency of historical use. Since all documents are checked out for use in the Archivo, there are historical complete usage records. About 10% of the manuscripts generate 40% of historical use.

Other criteria, however, were used to modify this main criterion, particularly that of completing a series even if that necessitated scanning less frequently used manuscripts. For example, all 300 bundles of the "Patronato" Series have been scanned. Conversely, the frequency of use criterion was also ignored at times: not all 5,000 bundles of the Contrataciones were scanned even though they were among the most frequently used, but only those that related to travel to the Americas. Other criteria included the state of conservation of the document (scanning those in the worst shape to avoid unnecessary future handling and to take advantage of the digital image enhancement capabilities), and ensuring adequate representation from all areas of the Americas.

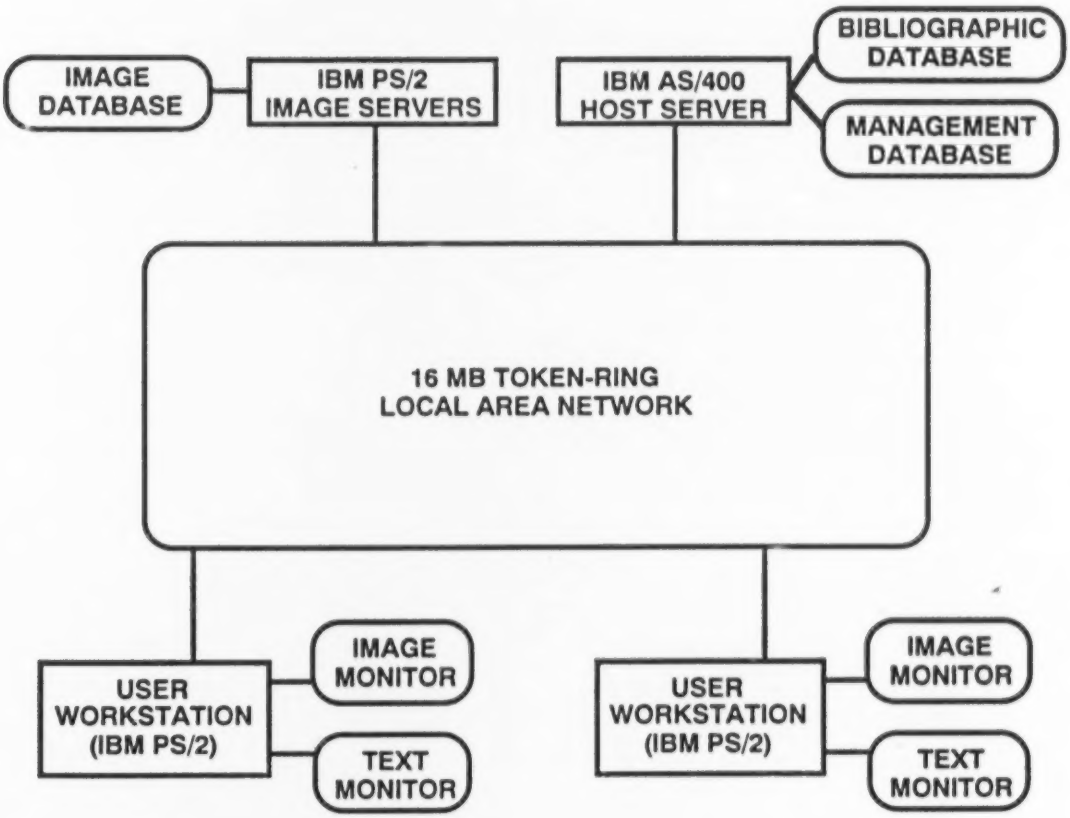
Other archives have been searched for documents pertaining to Spain's dominance of the Americas. For example, 500 bundles of documents in the archives of Simancas will be included in the project. This reflects the vision of the eighteenth-century King of Spain Carlos III to have all documents in one place, at the Casa Lonja in Seville.

TECHNICAL INFORMATION

At the time of this visit, Version 1, installed in 1988, of the overall software was being used by Archivo personnel in Seville. However, the demonstration and overview given in Madrid was of Version 2 of the software, which provides many enhancements. The plan was to install Version 2 in Seville a couple of months following the visit.

The system being developed consists of an Image Database, a Bibliographic Database, and a Management Database linked together by an IBM token-ring Local Area Network (LAN). These databases are accessible through workstations attached to the LAN located at various locations in the Archivo, most notably in an area set aside for researchers. This section reviews the various components of the project separately.

INSERT DIAGRAM HERE



Scanning

Scanning occurs in a workroom in the Archivo. There are 15 scanning stations at this time plus an additional two stations devoted to quality control. Xerox 7650 flatbed scanners attached to IBM PS/2's are used. Images are scanned at 100 dots per inch (dpi) — that is, at relatively low resolution — but at 16 grey levels (initially at 256 grey levels, but only the 16 most significant contiguous levels are retained), yielding an average of 1.4 MB (megabytes) per image.

These are subsequently compressed to about 350 KB (kilobytes) per image, using a compression algorithm tailored to the purpose by IBM Madrid Scientific center personnel adapted from the use of statistical coding (DCPM) compression techniques. The IBM personnel are also experimenting with a further compression refinement using an adaptive sampling scheme that tunes itself to the local characteristics of scanned pages: these typically yield a further 2-3 times factor in compression. The understanding, however, is that these latter techniques are not normally used in the project. This latter algorithm is quite fast, typically taking around 3-4 seconds to compress or decompress on an IBM PS/2 Model 80. The attitude seems to be that it is better to use an algorithm tailored to the purpose, rather than an emerging general standard such as JPEG or JBIG. IBM personnel, however, did indicate that they might well have adopted JPEG had it been around at the start of their project. There are no plans to convert to JPEG or JBIG, which is disappointing since the adoption of standards is to be preferred over minor savings in storage, particularly since the latter halve in cost every couple of years.

Images are not enhanced in any way at time of scanning. The philosophy used (which is believed correct) is to retain as much information as possible, deferring any image enhancement until actual time of use by the researcher. However, some enhancement does occur at time of scanning. Project personnel discovered, for example, that much of the "bleed-through" in the original documents is not captured if the scanner backing is black rather than white!

The compressed images are stored on Panasonic 9347 optical disks (or "flopticals") that have a maximum capacity per disk of about 900 MB. The understanding was that there are no plans at present to transfer these separate flopticals to any kind of "jukebox". Again, a proprietary image format, developed specifically for the project, is used to store the images, known as the "AG1" format (Archivo General 1), another concern from an international point of view. Since there are an average of about 1,800 pages in a bundle, an entire scanned bundle can be stored on a single optical disk (an "optical bundle," as it were), which is useful for indexing and retrieval purposes. This is an illusory benefit that will disappear when images are ultimately transferred to higher capacity/density storage, although there are no plans for such transfer at this time. (See comments on "Refreshing" below).

Scanning rates at each station average about 1 minute per page[1] when in production (of which 25 seconds is actual scanning), or about 350 pages per 7-hour day per workstation

---

[1] This can be compared, for example, with the production scanning rate of about 5 pages per minute attained with the Cornell/CPA/Xerox CLASS project.

allowing for rework, overhead, data entry, breaks, etc., or about 250,000 pages per month across all 15 scanning stations working two shifts per day (all numbers are rough estimates). Scanning personnel are provided by a subcontractor that charges 40 pesetas (about 40 cents US) per page[2].

These costs do not include preparatory work conducted by 15 archivists who prepare each bundle prior to scanning. Such preparation includes separating the manuscripts in the bundle as necessary (many are loosely sewn together); ordering them; creating an index document for subsequent entry into the Biblographic Database (see below), including the call number; and creating a sequential index that links the Bibliographic Database to the pages in the bundle. This information is written onto a floppy disk that is passed together with the bundle to the scanning station, together with a "control sheet" reflecting the contents of the floppy disk and including signatures tracking stages of progress through the process. The scanning operator uses a portion of this information as a scanning guide for control purposes to ensure that the right number of pages are scanned and in the right order.

Quality control is the responsibility of two stations devoted to post-scanning verification. At present, two people compare the contents of the floptical with the original scanning guide. This is a bottleneck at this time. It is hoped to automate part of the process. It is also planned to add another station.

Using this scanning technology, scanning the entire contents of the Archivo would require about 30 TB (terrabytes) of disk storage. The present project, expected to be completed in 1992, will require about 3 TB (or about 3,000 flopticals). About 80% of the scanning for this project had been completed by the time of this visit.

The above comments apply to grey-scale scanning, not to color images. There are about 8,000 color maps and prints in the Archivo, many of large format as large as 2 meters square, and there are plans to scan a number of these as part of the present project. The impression, however, is that these plans were still not firm at the time of the visit, and that most of the color scanning that had been accomplished was more testing and experimental for demonstration purposes, rather than production. Visitors were told that project staff were about to turn their attention more seriously to this phase of the project.

Several maps had been photographed onto Cibachrome film, and the film was scanned using a Nikon LS3500 scanner, 100 dpi, 8-bit color. Visitors saw some of these scanned images displayed on an IBM 6091 monitor (1,000 x 1,200 pixels), and they looked impressive, although some of the printouts provided were disappointing with loss of detail in the text annotations. Project personnel do not seem to be concentrating at this time on problems of loss of color fidelity among various transfer processes.

Image Database

The 3,000 or so flopticals will comprise the Image Database. These are kept in the computer room of the Archivo. An operator will manually mount a floptical on one of two

---

[2] Again, this is very high by CLASS standards, but can be accounted for by the relatively slow scanning rates.

PS/2 Model 80 servers (each server has several optical disk drives attached) upon a request generated by a researcher. The number of servers will be expanded to meet ultimate operational needs. The researcher will view both the page images and associated bibliographic information on a workstation linked to the servers across the local area network.

The system includes a caching strategy for speeding image transfers to the workstation. Decompression and image enhancement are performed at the researcher's workstation.

There are no plans at present to make the Image Database accessible over national or international networks. This is unfortunate, in spite of the proprietary nature of the image formats. Visitors were told that the bandwidth of Spain's networks (including IRIS, the network linking Spain's universities) is not adequate for this purpose. They are thinking of possibly publishing a CD-ROM version of the Image Database (or of selected portions) at some point in the future, both for general distribution and for location at other Spanish archives, but there are no firm plans. The focus of the project is on improved access by researchers who actually visit the Archivo.

There are also no firm plans for "refreshing" the Image Database to keep up with changing technology to take advantage of increasing storage capacities and lowering costs, and also to avoid technological obsolescence. The need to do so is recognized, however, and there is some hope that an endowment will be funded by the Foundation for this and other purposes. It seems that the Image Database is not backed up at this time: the hope had been that optical tape could be used for such purposes, but this has not proved possible. Making copies of the flopticals would be a cheap and effective form of backup.

Bibliographic Database

A separate computerized Bibliographic (or textual) Database is being constructed to the *entire* collection of the Archivo, not only the scanned portion. This is divided into two parts:

* An index to the 90% of the Archivo that are <u>not</u> being scanned into the Image Database. This will not be a new index, but a retrospective conversion of various catalogs and inventories dating back to the 18th century, comprising about 25,000 pages. The original descriptions will be modified to conform to a project standard (see below). At the time of my visit, 80% of this index had been converted and loaded onto the computer.

* A newly-constructed index to the Image Database itself with entries that contain more detail.

Both indexes are constructed and stored on an IBM AS/400 SQL database that is accessible over the LAN. It will occupy about 1 gigabyte of storage. Again, this is a proprietary approach (in spite of the use of SQL) because of the nature of the AS/400, although it would not take too much effort to convert the index to some less proprietary

format. This is also less of a problem than is the case with the Image Database, because there does seem to be some intent to make the Bibliographic Database accessible over both the Spanish academic network, IRIS, and, by extension, to international networks to which IRIS is linked. It is unclear, however, how this will work in practice. There are also plans to make the index available across the Ministry of Culture's own network.

The Bibliographic Database is full-text searchable, as well as accessible through the index structure itself (see Workstation Access, below).

Entries in both indexes will follow the "Method of Provenance," that is, a *hierarchical* index reflecting the original archival location of the bundle; the Section, such as the government department that originated the document; the Subsection, such as the responsible subdepartment; the Series; and other possible lower level information such as the pertinent country; or, in the case of Sailing Contracts, the port of destination; and finally the bundle itself and its call number.

Not all branches of the hierarchy follow the same exact pattern; neither are they of equal length. The substructure varies somewhat according to the particular material being classified, but the general approach is similar.

The index to the Image Database, however, will contain even more detail. The details captured are of two kinds: structural information reflecting the page by page structure of each bundle, and content information reflecting the actual contents of each page or sheet. These details vary by Series. For example, the index to Sailing Contracts includes such details as the name of the sailing vessel, the passenger's title and profession, any noteworthy identifying accomplishments, the name of the passenger's parents, and names of accompanying relatives and companions and of their parents.

Management Database

The Management Database also resides on the IBM AS/400. The purposes of the Management Database are to facilitate (i) researcher authorization for access to the Archivo, (ii) researcher access control to the Image and Bibliographic Databases, (iii) control of document movements to researchers and within and outside the Archivo, and (iv) the accumulation of system and usage statistics. A key objective is to provide logistical support to the Archivo Secretariat and the Chief of the Reading Room.

Researcher authorization support is provided through a system that records the accreditation of new researchers, adds them to the user file, and provides and modifies passwords for access to the various databases. Authorization entries refer to specific time periods, allowing for either temporary or permanent access.

Access control to the databases is provided through the password control system, with different levels of access provided to meet differing requirements (researcher, archivist, bibliographer, etc.). Furthermore, the system assigns physical workstation locations to researchers on each visit.

Document movements are also recorded and controlled. A user (researcher or other) requests a given bundle through the system. The document movement control module authorizes the request and issues appropriate document movement orders. Researchers may also reserve documents for use on specific days. At any time, the location of any bundle is

known to the system. An audit trail is also kept of who has accessed what documents, and when.

Usage and management statistics on all aspects of the system are accumulated and reported to appropriate levels of management.

Local Area Network

As already indicated, the system operates across a 16MB Token Ring Local Area Network that straddles the Archivo. Theoretically, this implies that about five compressed pages per second could be transmitted. In practice, this rate cannot be achieved because the LAN cannot operate at peak performance for sustained periods and because there are other limiting factors at the server and workstation ends.

Workstation Access

Access to all three databases is provided through various workstations. There will be about 60 workstations initially available to researchers located in a reading room of the Archivo. The workstations will mostly consist of IBM PS/2 Model 70's with two monitors attached: an IBM 8514 scrolling monitor for text display, and an IBM 8508 for grey-level display (1,200 x 1,600 pixels, 16 grey levels) of the scanned images. In some locations IBM 6091 monitors will be used to display scanned color images. These particular devices are, of course, subject to change depending upon what IBM equipment is available. The list price of a typical monochrome workstation is expected to be about $5,500.

There are different interfaces for access to the different databases. The interface for access to the Management Database seems fairly typical.

The interface to the Image Database is outstanding. It is designed primarily to enable researchers to display selected documents and to scroll or otherwise navigate through a bundle of documents stored on the floptical previously mounted on the image server using the scanning control information for referencing purposes; and to provide researchers with a set of computational tools for enhancing sections of the displayed images in real-time -- tools that are straightforward to use. These enhancement tools use adaptive and other filtering techniques for increasing contrast, and for removing document stains and ink bleedthrough. Palettes are provided for different kinds of transformations: linear, log, exponential, or customized. The tools have been carefully tailored to the particular characteristics of these documents, taking into account their reflectance and optical contrast, and particular types of artifacts encountered. In this case, such tailoring is appropriate since it occurs at the end user's workstation, not at the image server. Tools are also provided to select particular areas of the document for enhancement (including the ability, for example, to enhance the background of the document only without affecting the text), and to apply simple transformations to facilitate viewing such as inversion, rotation, or scaling.

The speed and ease of use of these tools are impressive. There is something almost magical in seeing a badly stained section of a 300-year old manuscript cleaned up before one's eyes and become legible again.

The interface to the Bibliographic Database is less impressive, and, by comparison, appears somewhat awkward to use. Even the developers had difficulty using it to navigate through the database. It is a somewhat limited textual approach to navigation through and around a rather straightforward hierarchical database. It lacks features and aids that can be provided by exploiting the strengths of graphical user interfaces. Nevertheless, it provides the kinds of capabilities one would expect for search and retrieval such as navigating up and down the hierarchy, retrieving by Boolean combinations of index terms, or linking to records related to a given field. Searching and retrieving appear to be not particularly fast, but this may be a characteristic only of the prototype shown. We expect there will be a need to improve this interface over time as researchers begin to use the system and offer feedback, to make it comparable in quality to the interface to the Image Database. There are a number of help tools provided, such as a built-in thesaurus, and a dictionary to aid in spelling conversions.

Printing

The system provides printing services for users, such as printing selected portions of the Bibliographic Database. It is also possible to obtain laser printouts of the enhanced screen images, or of an original unenhanced image stored in the Image Database.

CONCLUSIONS AND REMAINING ISSUES

This project shows what can be accomplished when funding and commitment co-exist. It is expected that there will be international promotion after the project's inauguration this spring: leaders are interested in widely publicizing the Archivo after remaining technical problems are solved. A 20-minute video about the project is expected to be available shortly, and a transportable prototype demonstration project is planned. A demonstration is planned for this year at the Huntington Library, San Marino, California. A selection of the digital image archive will be presented, and it is hoped that bibliographic searching and access will be accomplished over international networks to Spain.

There are problems, to be sure: the avoidance of open standards, local access only (at least to the Image Database), the lack of specific plans for the future, including plans for technology "refreshment." But these pale in comparison with the strengths.

Other issues still to be addressed:

-- As mentioned, the aspects related to printing hard copy (off microfiche and off the image database) are not completely worked out and require follow-up.

-- The entire system is conceived principally as a regional storage and access environment (with the exception of the bibliographic database, which will be available on an inter-Spain network).

-- The biggest remaining problem may be the management on new media of a massive amount of information. The Image Database from the first phase alone will consist of 3,000 optical disks. It remains to be seen whether the means of providing operational access by 30-35 simultaneous researchers will be adequate.

-- Although there was interest in further dissemination of the Bibliographic Database (through bibliographic utilities in other countries) and the Image Database (by making available subsets of optical disks to other archives and libraries), these aspects are not as yet at the top of the project's agenda. Discussion with the project's leadership concerning increased dissemination of the databases in various forms needs to continue. In particular, it would be useful to explore specific means for future cooperation and wider dissemination of the scanned materials.

In conclusion, as a large-scale reformatting project addressing an entire range of new problems, the work in Seville deserves continued attention. By any measure, this is an extraordinarily impressive digital scanning project, unmatched in scale and completeness. The methodology is best suited to older archival manuscripts rather than to book preservation. Nevertheless, there is much to learn for other applications. There is noteworthy commitment by all project sponsors to the success of the project, and an unspoken — but quite apparent — desire to extend the project to cover 100% of the Archivo General de Indias, as well as to other Spanish archives.

APPENDIX I: RECORD STRUCTURE, BIBLIOGRAPHIC DATABASE. "Apéndice B. Relación de datos por tipo."

Apéndice B. Relación de datos por tipo.

Información de control.

#000 Control actualización

Información básica.

#001 Tipo de entrada

#002 Encabezamiento

#003 Fechas extremas

#006 Signatura

#007 Incluido en signatura

#104 Condiciones de servicio

#014 Niveles de privacidad

#035 Clave autor responsable información básica

Información descriptiva.

#017 Contenido

#019 Clave fuente de información

#008 Signatura de procedencia

#011 Estado de conservación

#012 Sistema de ingreso

#013 Número de unidades

#025 Lugar de emisión

#026 Características internas

#027 Características externas

#028 Bibliografía de referencia

#029 Titulo propio

#030 Otros titulos

#031 Datos del autor

#032 Datos de la publicación

#033 Datos matemáticos

#034 Documentación aneja

#055 Notas

#056 Edición

Referencias de localización.

#020 Descriptor o relación especifica

Fechas para acotación.

#004 Fechas para acotación

Signaturas en otros soportes.

#010 Signatura en otro soporte

Signaturas antiguas.

#009 Signatura antigua

APPENDIX II: EXAMPLES OF FORMS COMPLETED BY ARCHIVISTS. The first is a bibliographic entry form for the Contrataciones Series. The second is the form completed for document control purposes that is passed to the scanning technicians.

SECCION: CONTRATACION
LEGAJO: 5472

Num.	Niv.	SIGNATURAS	POR	MOD
1	1	N.1		
2	2	R.1	1	
3	1	N.2		
4	2	R.2	1	R101

Preparado por: Diego de la Prada Espina. Fecha:
Revisado por: Fecha:
Digitalizado por: Fecha:
Verificado por: Fecha:

Número de serie de discos ópticos:

Anotaciones:

Incidencias:

LEG: CONTRATACION, 5530	N.	R	DESTINO:	FECHA:
	1	3	Veracruz	1786 06 20

PASAJERO	HIJO DE
SOLIS Y GOROSPE, José Mariano	

N V	LUGAR	TIT: Doctor
N		CARGO: Canónigo de merced de la Catedral de Antequera de Oaxaca.
		PROF: Clérigo

RELACION\ACOMPAÑANTES	HIJO DE
LÓPEZ FERNÁNDEZ, Ramón. Criado	LÓPEZ, Juan

NOTAS:

LEG: CONTRATACION, 5530	N.	R	DESTINO:	FECHA:
	1	4	Veracruz	1786 06 23

PASAJERO	HIJO DE
ANDRÉS Y AGUIRRE, Ginés de	

N V	LUGAR	TIT:
		CARGO: Primer pintor de la Real Academia de San Carlos de Nueva España.
		PROF: Pintor

RELACION\ACOMPAÑANTES	HIJO DE
PRIETO, María Antonia. Mujer	
Dos hijas pequeñas	
Un sobrino	
SOLER, Rosa. Criada	

NOTAS: Los números 4 a 7 aparecen juntos como un mismo expediente, aunque con licencias separadas.